



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/629,175	07/31/2000	Ophir Frieder	7519-164345	4562
7590	09/09/2005		EXAMINER	
Staas & Halsey LLP 1201 New York Avenue N W Suite 700 Washington, DC 20005			LE, UYEN T	
			ART UNIT	PAPER NUMBER
			2163	

DATE MAILED: 09/09/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	09/629,175	FRIEDER ET AL.	
	Examiner Uyen T. Le	Art Unit 2163	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) Responsive to communication(s) filed on 20 June 2005.
- 2a) This action is **FINAL**.                    2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) Claim(s) 1-33 and 44-62 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) 30-33,49 and 58-62 is/are allowed.
- 6) Claim(s) 1-29,44-48 and 50-57 is/are rejected.
- 7) Claim(s) \_\_\_\_\_ is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on \_\_\_\_\_ is/are: a) accepted or b) objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a) All    b) Some \* c) None of:
    1. Certified copies of the priority documents have been received.
    2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |   |   |
|---|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input checked="" type="checkbox"/> Interview Summary (PTO-413)          |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                    | Paper No(s)/Mail Date. _____.   |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____. | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
|   | 6) <input type="checkbox"/> Other: _____.                                   |

## **DETAILED ACTION**

### ***Response to Amendment***

Applicant is reminded to remove new matter introduced in the amendment filed 5 June 2003. Applicant added claim 43 reciting “semantic filtering” and amended the description of Figure 2 to support the claimed “semantic filtering”. The examiner requested removal of new matter in subsequent office actions.

Applicant’s arguments regarding amended claims have been fully considered but they are moot in view of the new grounds of rejection presented in this Office Action.

### ***Claim Objections***

Claim 45 is objected to under 37 CFR 1.75(c), as being of improper dependent form for failing to further limit the subject matter of a previous claim. Applicant is required to cancel the claim(s), or amend the claim(s) to place the claim(s) in proper dependent form, or rewrite the claim(s) in independent form. Claim 1 from which claim 45 depends already includes filtering based on parts of speech.

### ***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-29, 45-48, 50-57 are rejected under 35 U.S.C. 103(a) as being unpatentable over Aiken (US 6,240,409) of record.

Regarding claims 1, 45, Aiken discloses a method for detecting similar documents including all the claimed subject matter (see Figures 1a, b, column 3 lines 44-47). Note the step of obtaining a document 102, filtering the document 106. The claimed step of generating a tuple for the filtered document is met by the fact that a hash value and position pair is created and stored (see step 114, column 6, lines 7-28). The tuple is clearly compared with a plurality of tuples as claimed. Aiken discloses detecting if the document is similar to another document by determining if the tuple is clustered with another tuple in the document storage structured (see Figures 4a, 4b, 4c, column 7, lines 25-34, column 10, line 4- column 12, line 2). The claimed “tokens being eliminated based on parts of speech” is met by the fact that the method of Aiken eliminates stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3). Although Aiken does not specifically show sorting the filtered document to reorder the tokens according to a predetermined ranking, official notice is taken that it is well known in the art that different operating systems use different tokens ordering. Therefore, it would have been obvious to one of ordinary skill in the art to include sorting the filtered document to reorder the tokens according to a predetermined ranking in order to accommodate different operating systems while implementing the method of Aiken.

Regarding claim 2, Aiken discloses parsing and filtering the document (see column 4, lines 54-67). Clearly the filtered document comprises a token stream of a plurality of tokens as claimed.

Regarding claim 3, Aiken discloses retaining a token according to at least a token threshold (see column 11, lines 15-30) and tokens frequently .

Regarding claim 4, Aiken discloses that the retained tokens are arranged in the token stream (see Figure 4a, step 404).

Regarding claim 5, Aiken discloses determining the hash value for the filtered document by processing individually each retained token in the token stream (see column 6, lines 7-28, column 9, lines 24-26).

Regarding claim 6, Aiken discloses determining a score for each token in the token stream and comparing the score for each token to a first token threshold (see column 11, lines 15-30). The token stream is clearly modified by removing each token having a score not satisfying the first token threshold and retaining each token having a score satisfying the first token threshold as claimed since the document not containing a certain match ratio is discarded in the method of Aiken.

Regarding claim 7, although Aiken does not specifically show the step of comparing the score for each retained token to a second token threshold and modifying the token stream as claimed, Aiken explicitly show that not every substring's hash value is stored (see column 6, lines 29-30). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed feature while implementing the method taught by Aiken in order to further filter the document and save memory.

Regarding claim 8, Aiken discloses filtering by removing from the token stream at least one token corresponding to a stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 9, although Aiken does not explicitly disclose filtering by removing a duplicate of another token in the token stream, it would have been obvious

to one of ordinary skill in the art to include such a feature in order to avoid processing redundant token, thus saving time and resources.

Regarding claim 10, Aiken discloses removing a token from a token stream if the token is a very frequent token when Aiken shows that the method remove words of "the" "and" , "this" , "is" (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 11, Aiken discloses removing a token from a token stream (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 12, Aiken discloses removing formatting from the document (see column 4, lines 55-57).

Regarding claims 13, 14, clearly the method of Aiken uses collection statistics pertaining to a plurality of documents for filtering the document since the input file is compared to a set of collected files to detect similarity (see column 2, lines 47-51). The collection statistics have to be present for the collected documents to be clustered as shown in the method of Aiken (see Figure 4c, column 11, line 47- column 12, line 2).

Regarding claims 15-18, although Aiken does not explicitly show that the method uses specific hash algorithms as claimed, it is notoriously well known in the art to use different hash algorithms depending on users' requirements. Therefore, it would have been obvious to one of ordinary skill in the art to include all the claimed features while implementing the method of Aiken in order to suit users' needs.

Regarding claim 19, Aiken discloses a hash table (see column 12, lines 40-44).

Regarding claim 20, Aiken discloses that the document storage structure comprises a tree (see column 8, lines 30-38).

Regarding claims 21, 22, Aiken discloses that the tree comprises a binary tree (see column 8, lines 36-38). Although Aiken does not explicitly show that the binary tree is balanced, it would have been obvious to one of ordinary skill in the art to include such a feature in order to store data efficiently and to facilitate searching and localization.

Regarding claim 23, Aiken discloses a hash table and at least one tree (see column 5, lines 33-40, column 8, lines 30-38).

Regarding claim 24, Aiken discloses inserting the tuple into the document storage structure (see Figure 1a, 1b, 4a, 4b, 4c).

Regarding claim 25, the hash table of Aiken clearly comprises a plurality of bins of tuples as claimed and the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple at a bin of a hash table (see Figures 1, 2, 4c, column 7, line 46- column 8, line 33).

Regarding claim 26, Aiken discloses a tree comprising a plurality of branches, each bucket of the tree comprising at least one tuple and wherein the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple in a bucket of the tree (see column 8, lines 31-54, Figure 4c).

Claims 27, 29 correspond to a system to perform the method of claim 1, thus are rejected for the same reasons stated in claim 1 above.

Claim 28 corresponds to a computer program product to perform the method of claim 1, thus is rejected for the same reasons stated in claim 1 above.

Regarding claim 46, Aiken discloses removing frequently occurring terms (see column 4, lines 48-53).

Regarding claims 47-48, although Aiken does not specifically show removing infrequently occurring terms or words having an occurrence frequency that falls within a pre-determined frequency range, since users requirements vary, it would have been obvious to one of ordinary skill in the art to include the claimed features in order to accommodate users applications.

Regarding claim 49, although Aiken does not specifically show Unicode ordering, since Unicode is a recognized standard, it would have been obvious to one of ordinary skill in the art to include such ordering in order to use a standardized technique while implementing the method of Aiken.

Claim 50 recites the limitations of claim 1 without the sorting step, thus is broader than claim 1 and is rejected for the same reasons stated in claim 1 above.

Claim 51 corresponds to a system for claim 50, thus is rejected for the same reasons stated in claim 50.

Regarding claims 52-57, the claimed criteria for determining threshold and frequency scores merely read on notoriously well-known decision making techniques in the art. Therefore, it would have been obvious to one of ordinary skill in the art to include any criteria deemed appropriate while implementing the method of Aiken depending on users requirements.

Claim 44 is rejected under 35 U.S.C. 103(a) as being unpatentable over Aiken (US 6,240,409) of record, further in view of Haber et al (US 5,136,646) of record.

Regarding claim 44, Aiken discloses determining a hash value for a document (see Figure 1, column 4, line 17- column 7, line 45, column 9, lines 16-30), accessing a document storage structure comprising a plurality of hash values, each hash value representing one of a plurality of documents (see Figure 4a, column 10, line 4- column 11, line 46), determining if the hash value is equivalent to another hash value in the document storage structure (see Figure 4c, column 11, line 47- column 12, line 2). Although Aiken does not specifically show each tuple comprises a document identifier and a single hash value, it is well known in the art to hash a document into a single hash value as shown by Haber (see the abstract). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Aiken in order to detect document similarity instead of just portions of a document.

### ***Conclusion***

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Uyen T. Le whose telephone number is 571-272-4021. The examiner can normally be reached on M-F 7:00-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Safet Metjahic can be reached on 571-272-4023. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



UYEN LE  
PRIMARY EXAMINER

6 September 2005